Feature Selection for Rhizosphere Microbiome Studies in Presence of Confounding Using Standardization

Peng Liu

Iowa State University
E-mail: pliu@iastate.edu

Abstract: Microbiome studies have uncovered associations between microbes and plant, animal, and human health outcomes. This has led to an interest in identifying microbial interventions for treatment of disease and optimization of crop yields which will require the identification of individual relevant microbiome features. That task is challenging because of the high dimensionality of microbiome data and the confounding that results from the complex and dynamic interactions among host, environment, and microbiome. The performance of variable selection and estimation procedures may be unsatisfactory when there are differentially abundant features resulting from a categorical confounding variable. For microbiome studies with such a confounding structure, we propose a standardization approach to estimation of population effects of individual microbiome features. Due to the high dimensionality and confounding-induced correlation between features, we propose feature screening, selection, and estimation conditional on each stratum of the confounder. Comprehensive simulation studies are used to demonstrate the advantages of our approach in recovering relevant features. Utilizing a potential-outcomes framework, we outline assumptions required to ascribe causal, rather than associational, interpretations to the identified microbiome effects. We conducted an agricultural study of the rhizosphere microbiome of sorghum in which nitrogen fertilizer application is a confounding variable. We applied the proposed approach and identified microbial taxa that are consistent with biological understanding of potential plant-microbe interactions.